

# LD4L Data Archive

These RDF files were produced during the LD4L project. They are available for download as described below.

- [Converter output](#)
- [Usage data](#)
- [Additional triples](#)
- [Search](#)

## Converter output

The MARC records from each library were converted to BIBFRAME 1.0 RDF by the [Library of Congress mar2bibframe converter](#). LD4L's [bib2lod converter](#) was then used to produce RDF in the LD4L data model. The result is RDF in the [N-Triples](#) format. These dumps are available:

- [cornell.ld4l.full-catalog.2016-03-17.tar.gz](#) -- Converted triples for the Cornell catalog -- 13GB
- [harvard.ld4l.full-catalog-1.2016-03-24.tar.gz](#) -- Converted triples for the Harvard catalog (1 of 4) -- 5.4GB
- [harvard.ld4l.full-catalog-2.2016-03-22.tar.gz](#) -- Converted triples for the Harvard catalog (2 of 4) -- 5.2GB
- [harvard.ld4l.full-catalog-3.2016-03-21.tar.gz](#) -- Converted triples for the Harvard catalog (3 of 4) -- 6.3GB
- [harvard.ld4l.full-catalog-4.2016-03-22.tar.gz](#) -- Converted triples for the Harvard catalog (4 of 4) -- 5.4GB
- [stanford.ld4l.full-catalog-1.2016-03-23.tar.gz](#) -- Converted triples for the Stanford catalog (1 of 4) -- 3.4GB
- [stanford.ld4l.full-catalog-2.2016-03-22.tar.gz](#) -- Converted triples for the Stanford catalog (2 of 4) -- 3.4GB
- [stanford.ld4l.full-catalog-3.2016-03-21.tar.gz](#) -- Converted triples for the Stanford catalog (3 of 4) -- 4.0GB
- [stanford.ld4l.full-catalog-4.2016-03-22.tar.gz](#) -- Converted triples for the Stanford catalog (4 of 4) -- 4.8GB

## Usage data

StackScore usage data is available for the Cornell and Harvard holdings. The scores appear as annotations on the individual bib\_ids. Each file contains the usage data for the corresponding, similarly named file of converter output. Data is in N-Triples format.

These data files are available:

- [2016-03-17\\_cornell\\_ld4l\\_full\\_catalog\\_anno.tar](#) -- Usage data for the Cornell catalog -- 478MB
- [harvard.ld4l.full-catalog-1.2016-03-24\\_anno.tar](#) -- Usage data for the Harvard catalog (1 of 4) -- 286MB
- [harvard.ld4l.full-catalog-2.2016-03-22\\_anno.tar](#) -- Usage data for the Harvard catalog (2 of 4) -- 272MB
- [harvard.ld4l.full-catalog-3.2016-03-21\\_anno.tar](#) -- Usage data for the Harvard catalog (3 of 4) -- 296MB
- [harvard.ld4l.full-catalog-4.2016-03-22\\_anno.tar](#) -- Usage data for the Harvard catalog (4 of 4) -- 239MB

## Additional triples

Additional triples were created to supplement the converter output, adding Work IDs to the Works, and creating links across institutions, between corresponding Works and Instances.

A concordance file was created, associating all known OCLC numbers with their corresponding Work IDs. This file was made with data extracted from a recent Research snapshot of WorldCat, and is structured as follows:

- Column 1: every OCLC number found in a record from both 001 and 019
- Column 2: the current OCLC number for the record, from 001
- Column 3: the current Work ID associated with the record

Fields are tab-delimited. For example:

|           |           |          |
|-----------|-----------|----------|
| 100000569 | 100000569 | 49300684 |
| 100000668 | 100000668 | 83546218 |
| 100000767 | 100000767 | 83546282 |

Using this concordance file, each work was assigned a Work ID, based on the OCLC number of its instances. For example:

```
<http://draft.ld4l.org/cornell/n556b336629626fa2>
  <http://www.w3.org/2000/01/rdf-schema#seeAlso>
    <http://worldcat.org/entity/work/id/57063107> .
```

Although the data from the three institutions were stored in three separate triple-stores, owl:sameAs statements were created where possible to link matching works or matching instances in the separate collections.

Instances with matching OCLC identifiers were linked with owl:sameAs, as were Works with matching Work IDs.

These files are available:

- [cornell\\_additional\\_triples.tar.gz](#) -- Additional triples for the Cornell catalog -- 228MB
- [harvard\\_additional\\_triples.tar.gz](#) -- Additional triples for the Harvard catalog -- 217MB
- [stanford\\_additional\\_triples.tar.gz](#) -- Additional triples for the Stanford catalog -- 135MB

## Search

The project developed an experimental search service based on the converted data above. The application code is available from github ([https://github.com/ld4l/ld4l\\_blacklight\\_search](https://github.com/ld4l/ld4l_blacklight_search)) and is built on [Blacklight](#). Blacklight is a Rails app that includes a Solr search engine and the structure of the search index is determined both by [the Solr schema](#) and [the Blacklight catalog controller script](#).